AL/OE-TR-1993-0113

AD-A274 144

ARMSTRONG LABORATORY

EVALUATION OF QSAR FOR USE IN PREDICTIVE TOXICOLOGY MODELING

W.T. Brashear
P.P. Lu

MANTECH ENVIRONMENTAL TECHNOLOGY INC.
P.O. BOX 31009
DAYTON, OH 45437-0009

DTIC
ELECTE
DEC 2 7 1993
S E D

93-31106

JANUARY 1993

93 12 22 207

FINAL REPORT FOR PERIOD SEPTEMBER THROUGH DECEMBER 1992

Approved for public release; distribution is unlimited

AIR FORCE MATERIEL COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6573

NOTICES

When U.S. Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from the Armstrong Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center
Cameron Station
Alexandria, Virginia 22314

TECHNICAL REVIEW AND APPROVAL

AL/OE-TR-1993-0113

The experiments reported herein were conducted according to the "Guide for the Care and Use of Laboratory Animals," Institute of Laboratory Animal Resources, National Research Council.

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Terry A. Childress

TERRY A. CHILDRESS, Lt Col, USAF, BSC
Director, Toxicology Division
Armstrong Laboratory

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave Blank)

2. REPORT DATE
January 1993

3. REPORT TYPE AND DATES COVERED
Final Report, September-December 1992

4. TITLE AND SUBTITLE
Evaluation of QSAR for Use in Predictive Toxicology Modeling

5. FUNDING NUMBERS
Contract F33615-90-C-0532
PE 62202F
PR 6302
TA 630202
WU 63020223

6. AUTHOR(S)
W.T. Brashear, P.P. Lu

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
ManTech Environmental Technology, Inc.
P.O. Box 31009
Dayton, OH 45437-0009

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AL/OET
Armstrong Laboratory
Wright-Patterson AFB, OH 45433-6573

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AL/OE-TR-1993-0113

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

During the past decade, the field of Quantitative Structure-Activity Relationships (QSAR) has grown from the work of Corwin Hansch (Van Valkenburg, 1972) to other approaches that correlate structure with biological activity. The work of Hansch (Enslein and Borgstedt, 1989) correlated structure with activity for closely related agrochemicals and utilized physical parameters as QSAR descriptors. Recently, QSAR models for toxicological end points have been developed. Enslein developed an approach that used structural descriptors, connectivity indices, and shape indices as molecular descriptors (Enslein and Borgstedt, 1989). Klopman has pioneered artificial intelligence which uses computer generated substructural fragments as QSAR descriptors (Klopman, 1984). Computational chemists and molecular modelers have built a QSAR model with steric and electrostatic parameters generated from quantum mechanical calculations. Analyzing this three dimensional data by Comparative Molecular Field Analysis has been utilized to predict binding affinities (Cramer et al., 1988) It is also possible to combine these different approaches to generate new QSAR models.

In addition to the QSAR approach, there are expert-based systems or rule-based systems for correlating structure with activity. These computerized systems can apply heuristic rules from a knowledge base to a compound being queried. To accomplish this, a system must be able to recognize the structural features of chemical compounds. Expert systems have been written for the prediction of carcinogenicity, toxicity, and metabolism. An expert-based system is not a QSAR model, but it does offer the potential of making expert criteria for toxicological evaluations widely available.

14. SUBJECT TERMS
QSAR (Quantitative Structure-Activity Relationship)
computational chemistry
molecular modeling
predictive modeling of toxicological end points
predictive toxicology

15. NUMBER OF PAGES
19

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

PREFACE

The research reported herein was conducted by the Toxic Hazards Research Unit, ManTech Environmental Technology, Inc., and serves as a final report for the QSAR initial evaluation for use in predictive toxicology models. The research described in this report began in September 1992 and was completed in December 1992. It was performed under Department of the Air Force Contract No. F33615-90-C-0532 (Study No. F20). Lt Col James N. McDougal served as Contract Technical Monitor for the Toxicology Division, Occupational and Environmental Health Directorate, Armstrong Laboratory, Wright-Patterson Air Force Base, OH.

Accesion For	
NTIS CRA&I	X
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

DTIC QUALITY INSPECTED 2

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

BMDP	BMDP Software
CD ROM	Compact disc read only memory
CoMFA	Comparative Molecular Field Analysis
DAT	Digital audio tape
DEREK	Deductive Estimation of Risk from Existing Knowledge
EPA	Environmental Protection Agency
FDA	Food and Drug Administration
Gb	Gigabyte
HDi	Health Designs, Inc.
LHASA	Logic and heuristics applied to synthetic analysis
LC_{50}	Lethal concentration, 50%
LD_{50}	Lethal dose, 50%
LogP	Octanol-water partition coefficient
Mb	Megabyte
MOPAC	Molecular Orbital Package
PC	Personal computer
pK_a	Negative log of the acidity constant
QSAR	Quantitative Structure-Activity Relationships
RAM	Random access memory
SAS	SAS Institute
SAT	Structure Activity Team
SCSI	SCSI Device Interface
SPSS	Statistics Program for the Social Sciences
THRU	Toxic Hazards Research Unit
VAX/VMS	Digital Equipment, Inc. - Virtual Memory System

SECTION 1

INTRODUCTION

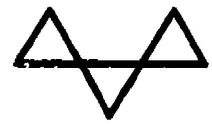
The Air Force has expressed a need for an alternative method for evaluating the toxicity of chemicals. Computational chemistry and Quantitative Structure–Activity Relationships (QSAR) offer a possible approach to the computer-assisted assessment of toxicity.

The objective of this study request was to evaluate software products that have the capability of estimating toxicological end points. The need for this type of capability at the Toxic Hazards Research Unit (THRU) is necessitated by the vast number of chemical substances that have no toxicological data. The toxicity of some of these compounds can be addressed through QSAR where a data base of structurally related compounds is available for comparison. The data base approach operates by correlating structural descriptors of an unknown with the descriptors of toxicologically characterized compounds contained in the data base. Another approach used by some software products is a rule-based system which has the ability to apply expert criteria to a compound. The rule-based, or expert system, applies a hierarchy of criteria to evaluate a toxicological end point.

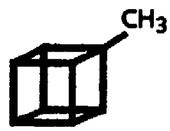
In cases where toxicological data from a data base or a rule-based system are not available, a QSAR model must be developed. The successful development of a QSAR model requires that appropriate molecular descriptors be included in the model. The molecular descriptors are then statistically correlated with experimentally obtained toxicological end points. When data from an appropriate number of compounds are available, a QSAR model may be used to evaluate the toxicity of unknown compounds. A set of compounds that would require the construction of a QSAR model are the high energy fuel additives shown in Figure 1.

These high energy fuel additives have little or no available toxicological data. In this case, it would be necessary to generate a set of toxicology data such as rat oral lethal dose, 50% (LD_{50}) values, or Ames mutagenicity tests. A QSAR model could then be developed to correlate levels of ring strain, bond energies, shape-dependent electrostatic forces, and the steric field of the molecules with the observed toxicity. This type of QSAR model development would require a computational chemistry program and appropriate statistical software.

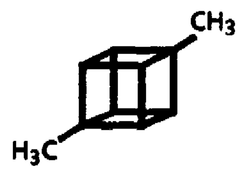
Spiropentane



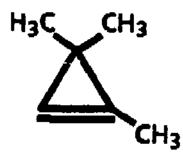
3-Triangulane



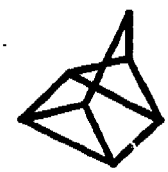
Methyl Cubane



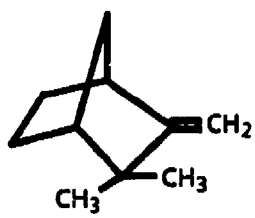
Dimethyl Cubane



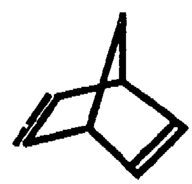
1,3,3-Trimethylcyclopropene



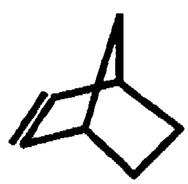
Quadricyclane



Camphene



Norbornadiene



Norbornylene

Figure 1. Compounds That Would Require the Construction of a QSAR Model.

SECTION 2

SURVEY OF COMPUTATIONAL CHEMISTRY PROGRAMS

Program: Topkat

Vendor: Health Designs, Inc.
183 East Main Street
Rochester, NY 14604
(716) 546-1464

Contact: Dr. Vijay Gombar

Topkat is a data base QSAR program that can predict a wide variety of toxicological end points. Among these are:

Carcinogenicity	Rat Oral LD_{50}
Mutagenicity (Ames)	Rat and Mouse Oral LD_{50}
Skin Irritancy (Draize)	*Daphnia magna* EC_{50}
Eye Irritancy (Draize)	Fathead Minnow LC_{50}
Mouse Inhalation LC_{50}	Aerobic Biodegradability
Rat Maximum Tolerated Dose	

The predictive capability of the Topkat program can be used to statistically estimate the toxicity of unknown chemical compounds. The Topkat program produces an estimate of a toxicological end point and statistical descriptors of the estimate. Among these are r^2 values, p values, F statistic, variance, and degrees of freedom. Each toxicological end point is estimated from its own data base, but each QSAR model used for Topkat is derived from a single heterogeneous data base (Enslein and Borgstedt, 1989). This is an advantage over a QSAR model, which is derived from smaller homologous data bases, and allows structurally diverse compounds to be searched using a single data base. The data bases used for the Topkat program are reviewed for accuracy, consistency, and methods of scoring. Data bases are cross-validated by removing observations one at a time, recalculating the QSAR model, and using the recalculated model to predict the toxicity of the removed observation. This method of validation tests how well a model predicts data rather than how well the model fits data.

The program has a user-friendly mode of operation and personal computer (PC) as well as VAX/VMS compatible versions. The Topkat program provides a graphical output identifying the structural features of a molecule which the QSAR algorithm has attributed to the toxicological prediction. The individual contribution of each structural feature is shown with the total estimate of toxicity. This allows the user to see the QSAR's association of structures with toxicological activity. If the confidence level of a toxicological estimate is low, the user is informed of this condition. This feature reduces the risk of extrapolating beyond the limit of the QSAR model.

The Topkat program provides validation of the toxicity estimate that it generates. This is accomplished by examining the compounds in the data base that were used to obtain the toxicological estimate. The structure of these compounds and their toxicities are displayed along with the functional groups and the specific topologies used by the QSAR model. This feature allows the operator to visualize how a QSAR result has been determined.

One disadvantage of this program is that the Topkat data bases are closed (Enslein, 1988). The addition of new compounds to the data base would require reparameterization of the QSAR model, and the present system does not have the ability to add compounds to an existing data base or to generate a new data base. In the future (1 year), Health Designs, Inc. (HDi) does intend to market a software product called Prognosys, that has the capability to generate QSAR parameters for a statistic software package such as SAS, BMDP, or Statistics Program for the Social Sciences (SPSS). When these parameters have been generated, they can be installed into a Topkat data base. This capability is necessary to generate and use QSAR for chemical compounds that do not fit the descriptors of an existing Topkat model. Such a set of chemical compounds are shown in Figure 1.

Cost:

Base PC Interface	$10,000
Carcinogenesis Module	$ 9,000
Rat Oral LD_{50}	$ 9,000

Terms:

These prices are for a permanent license. An additional annual maintenance fee needs to be paid after the first year. The annual maintenance fee is 15% of the current permanent license cost.

Terms vary with the number of users, the type of licensing agreement, and the number of prediction modules. A full system with all of the above listed modules would be about $80,000.

Hardware Requirement: IBM PC 80286, 80386, or 80486

VAX/VMS version of software available

Program: CaseTox II
Discovery Software, Inc.

Contact: Dr. Giles Klopman
Case Western Reserve University
Department of Chemistry
Cleveland, OH 44106
(216) 368-2000

CaseTox II is a QSAR program that can predict a variety of toxicological end points. Among these are: rat oral LD_{50}, carcinogenicity, teratogenicity, and other toxicological end points. The CaseTox II program uses substructural units as descriptors (Klopman, 1984).

QSAR models use multivariate linear regression, partial least squares, and discriminant analysis. These methods are used to analyze the QSAR data for a linear relation between the biological activity and the structural and/or chemical descriptors. The descriptors can be reactivity indices, structural parameters, molecular shape indices, partition coefficients, or other data generated from quantum mechanical calculations. Many QSAR calculations operate on closed data sets with parameters generated by the descriptors of the QSAR model. The CaseTox II program is different because it does not use a closed statistical method. The CaseTox II program generates a set of all possible substructural fragments in a data base and uses artificial intelligence to find appropriate descriptors. Many QSAR programs use preselected substructural keys, but the CaseTox II program employs an open-ended approach. Here, potential descriptors are evaluated through discriminant analysis and selected if they correlate with an observed property. The descriptors are utilized to form an open-ended set of keys which are used to evaluate biological activity. In the CaseTox II program, the QSAR model has the ability to learn from new compounds that can be added to a data set. The program is not tied to a predetermined set of QSAR parameters (e.g., the Topkat program by HDi).

The CaseTox II program can identify substructural fragments that are associated with biological activity (biophores) and fragments that are not associated with biological activity (biophobes). Compounds that do not contain a known biophore are assumed to be toxicologically inactive. The connectivity and the topology of the biophores is used to construct a QSAR model for estimating biological end points. The CaseTox II program outputs a probability-based estimate of the toxicity of the compound in question.

Because the open set of biophores and biophobes QSAR model works best on homologous or closely related data sets (Enslein,1988), the CaseTox II program does not use large heterogeneous data sets. The data bases used for the CaseTox II program tend to be small, and numerous data bases are needed to cover a wide range of chemical compounds. The CaseTox II program also has a module that can postulate metabolites of a chemical compound and the types of tissue that may produce them. The metabolites can then be searched for their own toxic effects.

The CaseTox II program does not appear to be for sale on an outright basis. The actual ownership of the program has been moved away from the original developers and the primary mode of availability appears to be to log on to their VAX system in Cleveland and to pay for the use of the program on a fee-for-use basis.

Cost per compound searched: $100

Cost per data base searched: $ 50

The advantage to this arrangement is that it is possible to utilize the program without a major capital expense.

Programs: **CompuDrug Expert Based Systems**

Vendor: CompuDrug North America, Inc.
P.O. Box 23196
332 Jefferson Road
Rochester, NY 14692-3196
(716) 292-6834

Contact: Dr. Harold Borgstedt

CompuDrug offers a series of programs that are expert-based systems for evaluating toxicological end points, possible metabolites, and physical properties. An expert-based system is rule-based artificial intelligence. Some of the CompuDrug programs also utilize a data base for the generation of rules. The programs offered by CompuDrug cover a number of different areas; three of these programs are:

MetabolExpert

This program uses a knowledge base and a data base of metabolic trees to predict possible metabolites of chemical compounds by establishing a structure–metabolism relationship. A chemical compound is submitted to the MetabolExpert program by drawing the chemical structure. When a structure has been submitted, sites of possible metabolic transformation are identified. Metabolites can be generated from the potential sites of metabolic transformation and a species specific semi-quantitative metabolic transformation scheme can then be generated. The MetabolExpert knowledge base can be expanded by the user through the generation of lesson files. MetabolExpert uses a reasoning-by-analogy approach to evaluate similarities between new information and existing metabolic data. This additional information can be incorporated into the program for future metabolic predictions.

HazardExpert

The HazardExpert program predicts toxic effects of organic chemicals based on molecular structure. The program uses a substructure-based expert system to predict toxicity. The program utilizes a knowledge base and rules for metabolic transformations. Bioavailability, bioaccumulation, and metabolism are also taken into consideration. The log P (octanol-water partition coefficient) and pK_a (negative log of the acidity constant) values are taken into consideration in estimating bioavailability. Bioaccumulation is estimated by using log P values, degree of metabolism, and duration of exposure. With the HazardExpert program, it is possible for the user to add his or her own compounds to the data base and to construct an in-house data base. The results of the toxicity query lists the bioaccumulation and bioavailability of the compound, and qualitative estimates of various toxic end points. Among these are carcinogenic potential, mutagenic potential, teratogenic potential, and neurotoxic effects.

ProLogP

ProLogP calculates the log P of a chemical compound based on structure. The program can also utilize a data base of known log P values that can be entered by the user. This allows the generation and utilization of an open data base of specialized log P values.

The pricing of these software products has been estimated based on current prices and a 25% government discount. VAX versions of the MetabolExpert program offer increased graphics capability and enhanced flexibility in the use of files. The programs are guaranteed against functional deficiencies for 6 months. During this initial time period any problems will be serviced free of charge.

MetabolExpert PC Version: $7,350

MetabolExpert VAX Version: $21,000

HazardExpert PC Version: $7,350

ProLogP PC Version: $985

Hardware Requirement: IBM PC 80286, 80386, or 80486

VMS/VAX version of some programs are available

Program: OncoLogic

Vendor: LogiChem Inc.
P.O. Box 357
Boyertown, PA 19512
(215) 367-1636

Contact: Ira M. Litman

OncoLogic is a rule-based program that determines the likelihood of a chemical compound being carcinogenic. This program is unique in that it has been developed with the cooperation of the Structure Activity Team (SAT) at the Environmental Protection Agency (EPA). The SAT at the EPA is responsible for assessing whether a chemical compound is a potential carcinogen. This is part of the premanufacture notice process, and is a determining factor regarding whether a chemical compound needs to have a bioassay for carcinogenicity. The OncoLogic program has been designed to give the same evaluation as the SAT team at the EPA.

The rules used by the OncoLogic program are applicable to metals, metal-containing inorganic compounds, polymers, fibers, and some organic chemicals. The ability to predict carcinogenicity of physical substances is unique because other programs rely solely on chemical structures. To accomplish this, the program utilizes parameters such as the chemical composition, the molecular weight (for polymers), the aspect ratio (for fibers), and particle size. Some of these parameters are designed to assess the bioavailability of the substance being queried. The rule-based system of the OncoLogic program considers the bioavailability of a substance in estimating the carcinogenicity. The final output includes a justification report which cites the rules used to estimate the concern of carcinogenicity. One weakness of the OncoLogic program is that at present the organic compounds portion of the program can only estimate the carcinogenic concern for aromatic amines. This excludes many organic chemicals. However, the company plans to expand the program to include other classes of organic chemicals. This effort is currently in progress.

The pricing of the software is structured so that an annual maintenance fee is paid. This fee includes upgrades to the OncoLogic programs and also provides for software and technical support.

OncoLogic, Fibers, Metals, and Polymers Program: $ 4,[illegible]00

OncoLogic, Aromatic Amines: $ 4,800

Hardware Requirement: IBM PC 80286, 80386, or 80486

Programs: Sybyl and QSAR Modules

Vendor: Tripos Associates, Inc.
1699 S. Hanley Road
Suite 303
St. Louis, MO 63144
(314) 647-1099

Contact: Scott Hutton

Tripos Associates markets a series of software products that can be used for the generation of computational chemistry parameters and QSAR model development. The Sybyl/Base package is a high-resolution molecular graphics and computational chemistry package. The Sybyl/Base package provides a user-friendly graphical interface that can be used to sketch molecular structures. Geometries can be optimized using force field type molecular mechanics computations. This can be used to examine the three-dimensional size, shape, and van der Waals volume of a molecule. The same interface also can submit computations to Molecular Orbital Package (MOPAC). MOPAC results can be used to display visual representations of molecular charge distributions, molecular orbitals, and isopotential surfaces. Calculations of this type can be entered into a molecular spreadsheet and used for QSAR model development.

Sybyl can interface with its own QSAR modeling package. The QSAR modeling package can combine computationally derived parameters with experimentally obtained parameters in a molecular spreadsheet. The QSAR program can use Comparative Molecular Field Analysis (CoMFA) to identify molecular structural and electrostatic regions that significantly affect activity. This is done by using a "probe atom" to compute the steric and electrostatic fields occupied by a molecular structure (Cramer et al., 1988). The fields of different molecules and observed biological properties can be incorporated into a QSAR model. Areas of the molecular field that vary with the property being modeled can be identified and cross-validated. Regions of the molecules where steric effects (or electrostatic effects) increase or decrease biological activity can be graphically displayed.

The modeling capabilities offered by Sybyl and the QSAR package are unique. The QSAR model can draw upon data from both computational chemistry calculations and toxicology studies. This type of approach could be utilized for constructing a QSAR model for the strained ring compounds shown in Figure 1. In general, the use of computational chemistry data with experimental data could be a useful tool in constructing a QSAR data base for compounds that are not adequately described by an existing QSAR model.

The pricing of the software is structured so that after the first year an annual maintenance fee of $8600 is required. This fee includes upgrades to the Tripos programs and also provides for software and technical support. Because of the computationally intensive molecular orbital calculations and high-resolution graphics, this software requires its own Unix workstation. A Silicon Graphics Iris workstation would be suitable for this software. The initial cost of a single user licensed copy of the Sybyl and QSAR software is as follows:

Base SYBYL Computational Chemistry Package	
QSAR Optional Module	
Comparative Molecular Field Analysis Option (CoMFA)	$60,000
Silicon Graphics Workstation	
Model: INDIGO XS24Z	
1280x1024, 24 BIT Color Graphics	
16" Color Monitor	
1.2 Gb Fast SCSI Systems Disk	
32 Mb Total System RAM	
4mm DAT Tape Drive for Back-Ups, CD ROM Reader	$22,455
Total System Fee	$82,455
Less Government Software Discount	$12,000
Total System Fee Including Discount	$70,455

Hardware Requirement: Because of the demanding computational requirements of molecular orbital calculations and the high-resolution graphics display, a UNIX based workstation with high-resolution graphics capability is required.

Program: DEREK

Vendor: School of Chemistry
University of Leeds
Leeds LS2 9JT
Tel: 0532 336531
Fax: 0532 336565

Contact: Dr. Philip N. Judson
Tel: 0943 880241

The DEREK (Deductive Estimation of Risk from Existing Knowledge) program is a knowledge-base system that provides an estimate of the toxic effects of chemical compounds (Sanderson and Earnshaw, 1991). The program operates through an "inference engine" that identifies chemical substructures within a molecule and relates this to a knowledge base of toxicological rules. The inference engine used by DEREK is based on the LHASA (Logic and Heuristics Applied to Synthetic Analysis) chemical synthesis program. The LHASA Program was originally designed to aid organic chemists in the development of synthetic strategies. The LHASA project was started 20 years ago and is a nonprofit organization consisting mainly of universities and corporations. The program uses a retrosynthetic approach that recognizes patterns or functional groups in a compound to be synthesized. The LHASA program can accept chemical structures as graphical input.

The DEREK program analyzes structures by a knowledge base that uses LHASA's retrosynthetic approach for the identification of functional groups. However, the strategy and display of the LHASA program have been modified to be applicable to toxicological end points. When the structural components have been identified, they are applied to a rule base. There are two sets of rules used by the DEREK program. One set of about 50 rules has been compiled by Schering Agrochemicals Limited. The second set of about 30 rules has been implemented from the U.S. Food and Drug Administration (FDA) structural alerts for carcinogenicity. A qualitative report is generated describing the toxicologically significant functional groups and the predominant physiological effects. The basic objective of the DEREK program is to identify a potential toxophore. If no toxophore is identified, a "no comment" statement is returned. The cost for acquiring the use of this software has an initial licensing fee and an annual maintenance fee for subsequent years. Training is included in these costs.

DEREK Program (first year): £ 10,000

Subsequent years: £ 7,500

Hardware Requirement: VMS/VAX

SECTION 3

DISCUSSION

The software products that have been reviewed to date have significant capabilities as computational aids for estimating toxicological end points. However, it is clear that the needs of the THRU can only be met by software that has the ability to generate it's own QSAR data base and descriptors for novel chemical compounds. The Sybyl program with the QSAR module sold by Tripos Associates, Inc. has this capability. Another program that will have the ability to construct a QSAR data base for novel chemical compounds is Topkat. HDi, the company that markets the Topkat programs, is scheduled to release a program module called Prognosys in about one year. Prognosys will allow Topkat users to generate parameters for the development of QSAR models. The ability to generate a QSAR model for unique compounds is critical, and has been demonstrated by the inability of current QSAR programs to generate toxicological data for the high energy fuel additives shown in Figure 1. During the course of this study, these structures were submitted to Topkat, CaseTox II, CompuDrug, and DEREK for QSAR evaluation. None of these programs were able to predict any toxicological end points because the respective data base or knowledge base did not contain information applicable to the compounds in question.

In addition to being able to predict the toxicities of novel chemical compounds, a good QSAR data base program would be a useful resource for the THRU. It could be utilized as a computational resource for obtaining QSAR toxicity data, and would be useful for ranking relative toxicities of homologous compounds by QSAR. The toxicology data contained in the Topkat program has been carefully screened for validity and uniformity of scoring. This is an essential ingredient for constructing a valid QSAR data base. The Topkat program is widely used by industry and government agencies. However, it is important to emphasize that at present the Topkat program can only estimate the toxicity of chemical compounds that are included in the existing Topkat data sets.

Other software products also are worthy of consideration. Among these are the knowledge-base programs. The OncoLogic program by LogiChem has been designed to utilize the EPA rules for carcinogenicity. These are criteria established by the SAT at the EPA. The OncoLogic program considers the physical state and estimates the bioavailability of a chemical. It does not rely solely on the structure of the chemical substance. However, one significant disadvantage of this program is that the program does not include many organic compounds and at present can only assess the carcinogenicity of aromatic amines. The DEREK program is also another knowledge-base program for estimating toxicity. The rules used by the DEREK program have been developed by Schering Agrochemicals Ltd., a member of the LHASA group, and also from the U.S. FDA's structural alerts for carcinogenicity. These programs offer a computerized simulation of the expertise available from these respective knowledge bases and can provide useful information concerning known toxophores.

Molecular modeling software would be a useful computational chemistry capability. A molecular modeling package is essential for visualizing the size, shape, and charge distribution of molecules. This can be important for the development of techniques for the separation and analysis of toxic chemicals in biological samples. The Sybyl computational chemistry package offered by Tripos Associates, Inc. has the ability to interface with a QSAR module. This provides the ability of using computational parameters in a QSAR data base. The Sybyl software is unique in this respect because it can perform QSAR modeling on parameters derived from molecular orbital calculations as well as experimentally derived parameters. This kind of capability is very important for developing a QSAR model for compounds that do not have descriptors found in other QSAR data sets such as the high-energy fuel additives shown in Figure 1.

Computational chemistry programs can provide data such as partition coefficients and possible metabolites. CompuDrug has programs that can calculate estimates of partition coefficients, acidity constants, and predict possible metabolites using a knowledge base. The CaseTox II program also has the capability of predicting possible metabolites from a knowledge base.

To summarize, the best QSAR data base program appears to be Topkat by HDi. It is a well-constructed data base that will eventually have the capability to generate a QSAR model for novel chemical compounds. The best computational chemistry program is Sybyl with the QSAR module by Tripos Associates. This program will allow the development of a QSAR data base for novel chemical compounds. Table 1 categorizes the basic features of the programs reviewed. It should be pointed out that a simple classification of some of these programs is difficult. For instance, the CompuDrug programs are knowledge base programs that encompass an expandable data base. Programs that offer an expandable data base do so in a manner that requires a certain level of expertise by the user, and should not be thought of as a turn-key operation.

TABLE 1. SUMMARY OF QSAR PROGRAMS

Program	Data/Knowledge Base	Develop QSAR Model	Expand Data base	Toxicological End point
Topkat	Data base	Yes[1]	No	Diversified[2]
CaseTox II	Data base	No	Yes	Diversified[2]
CompuDrug	Knowledge base	No	Yes	Qualitative Description Toxic End points
OncoLogic	Knowwldege base	No	No	Carcinogenicity
Sybil	Data base	Yes	Yes	Defined by user
DEREK	Knowledge base	No	Yes	Qualitative Statement of Risk

1 Future capability to be introduced in 1993
2 A variety of different toxic end points such as rat oral LD_{50}, genotoxicity, carcinogenicity.

SECTION 4

REFERENCES

Cramer, R.D. III, D.E. Patterson, and J.D. Bunce. 1988. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier protein. *J. Am. Chem. Soc.* 110:5959-5967.

Enslein, K. 1988. An overview of structure-activity relationship as an alternative to testing in animals for carcinogenicity, mutagenicity, dermal and eye irritation, and acute oral toxicity. *Toxicol. Ind. Health* 4(4):479-498.

Enslein, K. and H.H. Borgstedt. 1989. A QSAR model for the estimation of carcinogenicity: example application to an azo-dye. *Toxicol. Lett.* 49:107-121.

Klopman, G. 1984. Artificial intelligence approach to structure-activity studies. Computer automated structure evaluation of biological activity of organic molecules. *J. Am. Chem. Soc.* 106:7315-7321.

Sanderson, D.M. and C.G. Earnshaw. 1991. Computer prediction of possible toxic action from chemical structure; the DEREK system. *Hum. Exp. Toxicol.* 10:261-273.

Van Valkenburg, J.W. 1972. *Biological Correlations – The Hansch Approach*, Washington DC: American Chemical Society.

U S.G.P.O.:1993-650-057/81029